# Different Approaches for Face Authentication as Part of a Multimodal Biometrics System

*Jaromir TOVAREK, Miroslav VOZNAK, Jan ROZHON,*
*Filip REZAC, Jakub SAFARIK, Pavol PARTILA*

Department of Telecommunications, Faculty of Electrical Engineering and Computer Science,
VSB–Technical University of Ostrava, 17. listopadu 15, 70833 Ostrava, Czech Republic

jaromir.tovarek@vsb.cz, miroslav.voznak@vsb.cz, jan.rozhon@vsb.cz,
filip.rezac@vsb.cz, jakub.safarik@vsb.cz, pavol.partila@vsb.cz

DOI: 10.15598/aeee.v16i1.2547

***Abstract.*** *This paper describes different approaches for the face authentication from the features and classification abilities point of view. Authors compare two types of features - Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) including their combination. These parameters are classified using Multilayer Neural Network (MLNN) and Support Vector Machines (SVM). Face authentication consists of several steps. The first step contains Viola-Jones algorithm for face detection. Authors resize the detected face for a fixed vector and afterwards, it is converted into grayscale. Next, feature extraction with a simple Min-Max normalization is applied. Obtained parameters are evaluated by classifiers and for each detected face, authors get posterior probability as the output of the classifier. Different approaches for face authentication are compared with each other using False Acceptance Rate (FAR), False Rejection Rate (FRR), Equal Error Rate (EER), Receiver Operating Characteristic (ROC) and Detection Error Tradeoff (DET) curves. The results are verified with AR Face Database and elaborated in a feature extraction and classifier design point of view. Best results were achieved by HOG feature for SVM classifier. Detailed results are listed in the text below.*

## Keywords

***Face authentication, HOG, LBP, MLNN, SVM.***

## 1. Introduction

Personal authentication can be divided into three fields according to methods used. The first field is based on knowledge, which means that the person knows a password. The second field is represented by the authentication methods based on possession (identification card, key). The last one is based on the biometric authentication. The systems coming from the biometric authentication are used to verify the identity of a person by using unique physiological features (fingerprint, iris, retina, facial geometry, voice, etc.)[1]. The main advantage of biometric authentication is that a user does not need to remember a password or always carry an easily stealable key. The reasons for using biometric authentication are speed, convenience, precision, high reliability, zero operating cost, practicality and clarity. Biometric authentication can be used in many areas: security of computers and data, building access, judiciary, ensuring a comfort, etc. [2] and [3].

Face recognition represents a technology which identifies and verifies a unique facial geometry from the digital image. Face recognition can be divided into two areas. The first area is the face identification and the second is face authentication [4]. Face recognition is widely used because the facial geometry is one of the very popular biometric characteristics. Digital image of a face can be scanned simply and non-invasively with common camera equipment. There are many areas where we can use face recognition (access control, bankcard identification, security monitoring, etc.) [5].

A lot of work has been done in the last years in the field of face authentication as part of a multimodal biometrics system. Sanderson et al. [6] provide a review of important milestones in audio-visual person identification and verification (features, classifiers and fusion techniques). Authors [6] used eigenface as features and GMM for classification in their research. Brunelli et al. [7] used a set of geometric features, describing the size and the layout of the different features in the faces

(eye, mouth, nose, eyebrow). Recognition proceeded by measuring the distance of the unknown descriptive vector and a set of reference vectors (known people). Raghavendra et al. [8] compared four methods for feature extraction (PCA, 2DPCA, LDA, 2DLDA). Each of these feature vectors was classified by nearest neighbour classifier. Kala et al. [9] used a set of geometric features (width of the eye, length of the eye, length of the mouth, width of the mouth, ...) for face representation. These parameters were classified by artificial neural network. Barbu et al. [12] used SIFT-based face recognition technique for feature extraction. Authors used measurement of the distance between feature vectors for classification. In [10] and [13], authors used the same features (HOG, LBP) as a descriptor of face. Chandrasheker et al. [10] used SVM and HMM for classification and Xie et al. [13] used only SVM.

This paper is focused on face authentication and compares relevant methods to achieve the lowest error rate. Authors compare various parameters (HOG, LBP and their combination) and multiple classifiers (MLNN and SVM). The combination of parameters and classifier with the lowest error rate will be used for multimodal biometrics system in future work. This multimodal system will consist of voice authentication and face authentication.

The rest of the paper is organised as follows: the second chapter mentions the basic idea of face authentication and is followed by the description of the AR Face Database that has been used in the presented experiment. The results are then presented in chapter four, and section five contains a discussion about the future work and possible improvements.

# 2. Face Authentication

Face authentication or face verification is a kind of biometric authentication where the facial geometry is used for the verification process. Simply put, the main task of the technology is to decide whether a face from the digital image belongs to an authenticated user or not. Face authentication is used in many areas such as banking, building access, devices access and so on. As already mentioned, the main advantages of this approach are low price, user comfort, contactless nature and sufficient accuracy [4] and [5]. Authors have focused on face authentication because the goal is to design multimodal biometric authentication system which will consist both of the face and voice authentications.

The process of authentication consists of following steps: face detection, preprocessing (resize, grayscale conversion), feature extraction, classification and decision. These steps are shown in Fig. 1 and described in more details below.

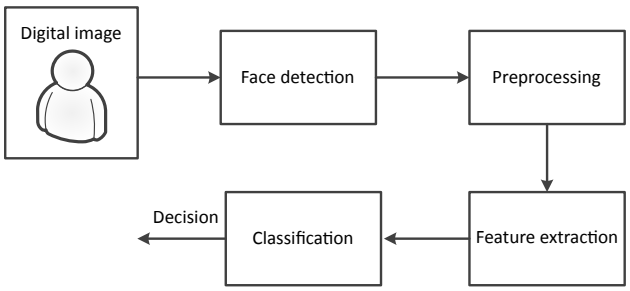



**Fig. 1:** Process of face authentication.

## 2.1. Face Detection

Face detection is intended to find a face and its coordinates in a given image. Authors used Viola-Jones detection method [14] in the presented experiment. This method is based on three main features (integral image, AdaBoost training, cascading classifiers). Viola-Jones algorithm is very fast, accurate and very suitable for face detection [15].

## 2.2. Preprocessing

Preprocessing performs an adjustment of a detected face into a useful form. It consists of two parts. The first part is the change of the size of the face picture. We have to resize a detected face for the classifiers (we need the same size of feature in all time). Authors have set up the size of the detected face to 120×120 pixels. In the second part, the detected face is converted into grayscale.

## 2.3. Feature Extraction

The most important step of the face authentication is the choice of significant parameters/features. These parameters should meet some requirements. First, they should be robust. Parameters should not change their characteristics in time. Second, they should be secure, which means that it should not easy to mimic these parameters. Third, they should be both illumination and rotation invariant [4] and [5]. The most used descriptors are HOG and LBP [11], [16] and [17] and these have also been used in the presented experiment.

### 1) Histogram of Oriented Gradients

The method is based on evaluating well-normalised local histograms of image gradient orientations in a dense grid. The basic idea is that local object appearance and shape can often be characterised rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge position. Computation algorithm is

described in [18]. We used these input arguments for HOG extraction: size of HOG cell was set up 8×8 pixels, number of cells in block was 4, number of overlapping cells between adjacent blocks was 1 and number of orientation histogram bins was set up 9. This setting corresponds to the length of HOG feature 7056 for image size 120×120 pixels.

#### 2) Local Binary Patterns

The basic LBP method characterises the spatial structure of a local image texture by thresholding 3×3 square neighbourhood with the value of the center pixel and considering only the sign information to form of a local binary pattern [17]. LBP is defined by Eq. (1).

$$LBP(x_c, y_c) = \sum_{u=0}^{U} s(I_u - I_c)2^u, \tag{1}$$

where $x_c$ and $y_c$ are coordinates of pixel, $I_c$ is a brightness level of center pixel, $I_u$ is a brightness level of neighboring pixel, $s(I_u - I_c)$ is the threshold function and $U$ is a number of neighboring pixels.

### 2.4. Classification

From the classifiers point of view, authors compare two types of classification methods. The first method is MLNN. This method has appropriate properties for face authentication (high accuracy, generalisation, adaptation) [5]. The second method is SVM. This method is very useful for face authentication because it is primarily intended for binary classification [4].

#### 1) Multilayer Neural Network

Authors have used feedforward Multilayer Neural Network with backpropagation in the experiment [19]. The network consisted of three layers (input, hidden and output layer). The number of neurons in the input layer is determined by the number of extracted parameters for a given detected face (for example 7056 HOG). The number of neurons in hidden layer was set up to 10. Output layer represented two output classes (reference user, imposter). The sigmoid was used as an activation function with the steepness of 0.5 for each neuron.

#### 2) Support Vector Machines

SVM offers a progressive method in the field of machine learning. The principle of classification is to find the hyperplane that divides the training data into the feature space. The optimal hyperplane is such that the training data points lie in the opposite half-space and the value of the distance between half-spaces is the largest. In other words, the goal is to maximise space among half-spaces (maximum margin). Support vectors are described by training data points that represent a decision-making role [4] and [5].

### 2.5. Decision

The last step of the authentication process is the decision about allowing the access or not. The system has to decide whether the user is the reference one or the imposter. The decision is based on comparison of max value of score and threshold. If the max value is higher than a threshold, the user is marked as reference one otherwise as an imposter.

Measurement of the face authentication performance allows comparison of different systems. Authors have used FAR, FRR, EER, ROC and DET curves for measurement of performance. The FAR is the measure of the likelihood that the face authentication system will incorrectly accept an access attempt by the imposter. FAR is computed by Eq. (2). The FRR is the measure of the likelihood that the face authentication system will incorrectly reject an access attempt by a reference user. FRR is computed by Eq. (3). EER indicates that the proportion of FAR is equal to the proportion of FRR. ROC shows the relationship between true positive rate (sensitivity) and False Positive Rate (FAR) at various threshold settings. DET curve is a graphic representation of error rates (FAR vs FRR) for binary classification systems [20].

$$FAR = \frac{N_{FA}}{N_{IVA}}, \tag{2}$$

where $N_{FA}$ is the number of incorrect acceptance and $N_{IVA}$ is the number of all imposter attempts.

$$FRR = \frac{N_{FR}}{N_{EVA}}, \tag{3}$$

where $N_{FR}$ is the number of incorrect rejection and $N_{EVA}$ is the number of all authorized attempts.

## 3. AR Face Database

The database contains over 4000 colour frontal view images of 126 people's faces (70 men and 56 women) that were taken during two different sessions separated by 14 days. Similar pictures were taken during the two sessions. No restrictions on clothing, eyeglasses, make-up, or hairstyle were imposed upon the participants. Controlled variations include facial expressions (neutral, smile, anger, and screaming), illumination (left

light on, right light on, all side lights on), and partial facial occlusions (sunglasses or a scarf) [21].

Authors have chosen 18 reference participants (13 men and 5 women) and 10 imposter participants (6 men and 4 women) for the experiment. It corresponds to the number of participants in the correlation speech database. For each person, a 16 digital images have been used for training and 10 for testing of the system.

## 4. Experimental Results

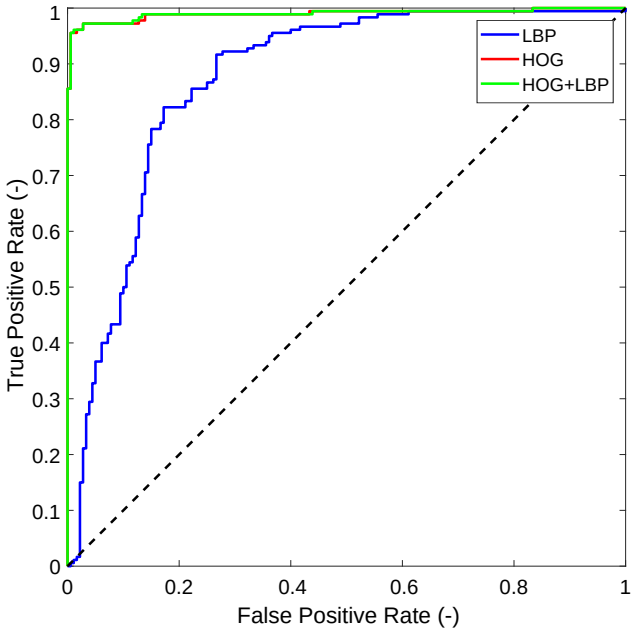
The SVM model and MLNN were trained for each of 18 reference users. These classifiers were used for recognition between two classes (class of authorised user and imposters class). Class of authorised user was trained using 16 digital images (this set of images contained images from both sessions), 10 remaining images were used for testing. The imposters class was trained as a background model. We used 1 digital image from each of 17 reference users for training this model. As the testing data for imposters, digital images from 10 participants (10 imposters) were used. These participants do not belong to the reference users (background model was not trained by using digital images from these participants). This training process was repeated for LBP only, HOG only and LBP+HOG features. The results for both classifiers with different features are shown in Tab. 1. The table contains the values of FAR and FRR with threshold 0.5 (50 %) and the value of EER. The values of these parameters are given in percent. ROC curves for both classifiers are shown in Fig. 2.

**Tab. 1:** Results for all features - SVM and MLNN classifiers (threshold 50 %).

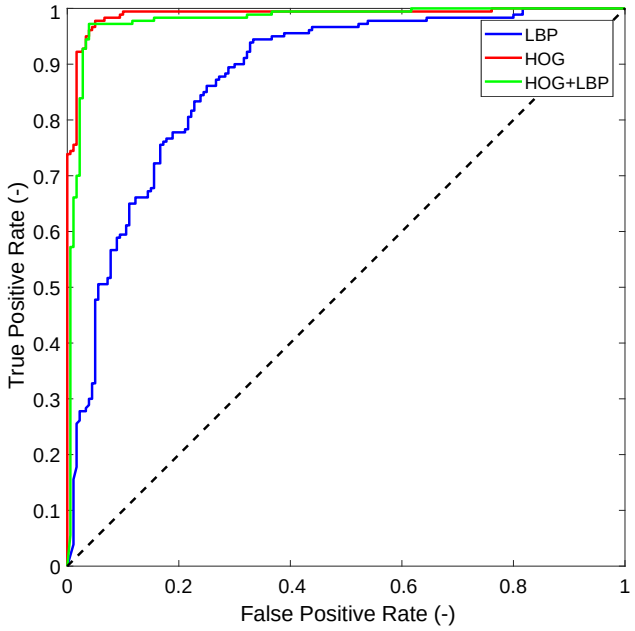
| Features | SVM | | | MLNN | | |
|---|---|---|---|---|---|---|
| | FAR | FRR | EER | FAR | FRR | EER |
| LBP | 35.0 | 6.6 | 17.8 | 26.6 | 13.9 | 21.7 |
| HOG | 2.7 | 3.3 | **2.8** | 6.0 | 2.2 | **3.9** |
| LBP + HOG | 2.7 | 3.3 | **2.8** | 7.2 | 2.7 | 3.9 |

As shown in the table above, best results were achieved by HOG features for SVM classifier. The values of FAR and FRR were 2.7 % and 3.3 % for threshold 50 %. It corresponds to accuracy 96.9 %. EER was 2.8 % for these parameters. The lowest EER (3.9 %) was achieved by using the same parameters for MLNN classifier. Combination of LBP and HOG parameters brings similar results as HOG parameters. From the classifier point of view, the SVM classifier achieved a better result for all features when compared to MLNN classifier. More detailed results are listed only for SVM classifier with HOG features.

From the authentication point of view, we want to achieve the lowest FAR. It means that the threshold has to be set up to a high value. Figure 3 shows the values of FAR and FRR depending on the threshold. The zero value of FAR was achieved after setting the threshold to 62 %. Table 2 shows confusion matrix for threshold equal 50 %. As we can see in Fig. 3 the value of FAR decreases with an increasing threshold. On the other hand, the value of FRR increasing as expected. Figure 4 shows DET curve with a marked point of EER.




(a) ROC curve - SVM classifier.




(b) ROC curve - MLNN classifier.

**Fig. 2:** ROC curves.

**Tab. 2:** Confusion matrix - SVM classifier, HOG features with threshold 50 %.

| | | | | |
|---|---|---|---|---|
| **Output classes** | *Reference users* | 174<br>48.3 % | 5<br>1.4 % | 97.2 % |
| | *Imposters* | 6<br>1.7% | 175<br>48.6 % | 96.6 % |
| | | 96.6 % | 97.2 % | **96.9** % |
| | | *Reference users* | *Imposters* | |
| | | **Target classes** | | |

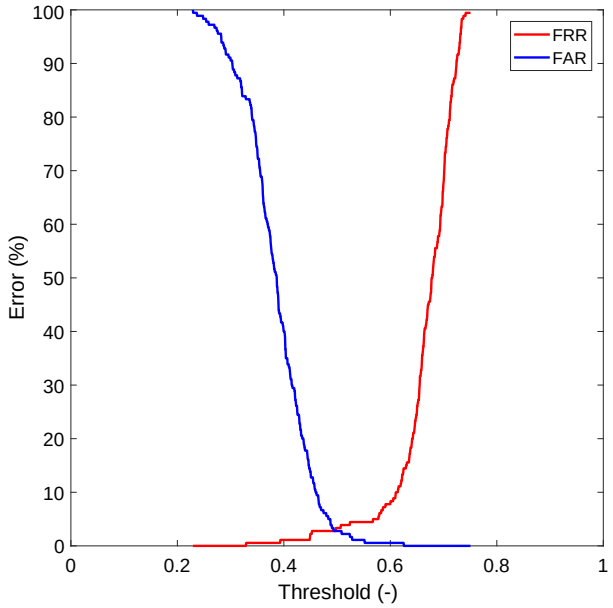



**Fig. 3:** FAR vs FRR graph.

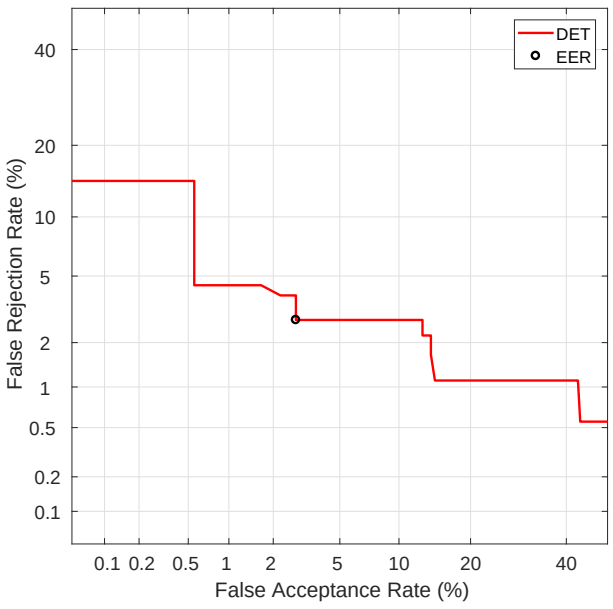



**Fig. 4:** Detection error tradeoff graph with EER.

# 5. Conclusion and Future Work

The aim of our research was to find the best features for face authentication and suitable classifier with the lowest values of FAR and FRR. The results and knowledge of this research will be used for the design of multimodal biometrics system based on voice and face authentication. Our research was focused on the analysis of AR Face Database using different parameters and classifiers. We compared LBP, HOG features and their combination. Authors used two machine learning methods for classification (MLNN and SVM). The comparison was made based on values FAR, FRR and EER.

From the classifiers point of view, we achieved lower values of FAR, FRR and EER with SVM classifier than with MLNN classifier. This result follows the properties of classifiers. The SVM classifier uses "only" support vectors for classification. It means SVM classifier does not need big training data set if the training data set contains suitable support vectors. On the other hand, MLNN classifier needs big training data set for precise neuron weights setting. Conclusion of this result is that if we have small training data set we should use SVM classifier. Experimental results show the best values of errors were achieved for HOG features. HOG performs better than LBP because while binary local pattern feature takes care of a local pattern, histogram of gradients, on the other hand, investigates the ensemble (histogram) of changes (gradients). Therefore, it is expected that investigating the whole image rather than looking for local patterns should perform better. Classification errors occurred primarily for images where a participant wears a scarf. It means that the values of FAR, FRR and EER will be lower when we do not use these images.

If we want to compare our results with other research works it is important to say that almost all of the works reviewed in the introduction used different databases and/or different experimental setup, thus any direct comparison between the numerical results would be meaningless. If we compare only suitability of classifiers we can say that SVM classifier is the most used classifier for face authentication. This fact corresponds to the results of experiments mentioned in the introduction and our results.

The future work will be divided into two parts. The first part will be focused on expanding our speech database (Comtech) with photos of faces. The second part will contain a design of the multimodal biometrics authentication system based on voice authentication and face authentication.

## Acknowledgment

This work was supported by the grant of Technology Agency of the Czech Republic reg. no. TF01000091 and partially by the SGS grant reg. no. SP2017/174 conducted at VSB–Technical University of Ostrava. This article was prepared within the frame of sustainability of the project No. CZ.1.07/2.3.00/20.0217 "The Development of Excellence of the Telecommunication Research Team in Relation to International Cooperation" within the frame of the operation programme "Education for competitiveness" that was financed by the Structural Funds and from the state budget of the Czech Republic.

## References

[1] DE LUIS-GARICA, R., C. ALBEROLA-LOPEZ, O. AGHZOUT and J. RUIZ-ALZOLA. Biometric identification systems. *Signal Processing*. 2003, vol. 83, iss. 12, pp. 2539–2557. ISSN 0165-1684. DOI: 10.1016/j.sigpro.2003.08.001.

[2] PATO, J. and L. MILLETT. *Biometric recognition: challenges and opprotunities*. 1st ed. Washington, D.C.: National Academies Press, 2010. ISBN 978-0-309-17704-7.

[3] POH, N. and M. SCHUCKERS. Biometrics statistics: a foreword and introduction to the special issue. *IET Biometrics*. 2015, vol. 4, iss. 4, pp. 206–208. ISSN 2047-4938. DOI: 10.1049/iet-bmt.2015.0100.

[4] TOLBA, S., A. EL-BAZ and A. EL-HARBY. Face recognition: A literature review. *International Journal of Signal Processing*. 2006, vol. 2, iss. 2, pp. 88–103. ISSN 2047-4938. DOI: 10.5120/ijais2016451597.

[5] ZHAO, W., R. CHELLAPPA, P. PHILLIPS and A. ROSENFELD. Face recognition. *ACM Computing Surveys*. 2003, vol. 35, iss. 4, pp. 399–458. ISSN 0360-0300. DOI: 10.1145/954339.954342.

[6] SANDERSON, C. and K. PALIWAL. Identity verification using speech and face information. *Digital Signal Proccesing*. 2004, vol. 14, iss. 5, pp. 449–480. ISSN 1051-2004. DOI: 10.1016/j.dsp.2004.05.001.

[7] BRUNELLI, R., D. FALAVIGNA, T. POGGIO and L. STRINGA. Automatic person recognition by acoustic and geometric features. *Machine Vision and Applications*. 1995, vol. 8, iss. 5, pp. 317–325. ISSN 1432-1769. DOI: 10.1007/BF01211493.

[8] RAGHAVENDRA, R., A. RAO and H. KUMAR. Multimodal person verification system using face and speech. *Procedia Computer Science*. 2010, vol. 2, iss. 1, pp. 181–187. ISSN 1877-0509. DOI: 10.1016/j.procs.2010.11.023.

[9] KALA, R., H. VAZIRANI, A. SHUKLA and R. TIWARI. Fusion of Speech and Face by Enhanced Modular Neural Network. In: *Proceedings of the Fourth International Conference on Information Systems, Technology and Management*. Heidelberg: Springer, 2010, pp. 363–372. ISBN 978-3-642-12035-0. DOI: 10.1007/978-3-642-12035-0_37.

[10] CHANDRASHEKER, T. R. and A. K. GAUTAM. Face Recognition based on Histogram of Oriented Gradients, Local Binary Pattern and SVM/HMM Classifiers. *International Journal of Engineering Sciences and Research Technology*. 2014, vol. 3, iss. 8, pp. 344–352. ISSN 2277-9655. DOI: 10.1.1.683.3019.

[11] GHORBANI, M., A. TARGHI and M. DEHSHIBI. HOG and LBP: Towards a robust face recognition system. In: *2015 Tenth International Conference on Digital Information Management (ICDIM)*. Jeju: IEEE, 2015, pp. 138–141. ISBN 971-4673-9152-8. DOI: 10.1109/ICDIM.2015.7381860.

[12] BARBU, T., A. CIOBANU and M. LUCA. Multimodal biometric authentication based on voice, face and iris. In: *2015 E-Health and Bioengineering Conference (EHB)*. Iasi: IEEE, 2015, pp. 1–4. ISBN 978-1-4673-7545-0. DOI: 10.1109/EHB.2015.7391373.

[13] XIE, Z., P. JIANG and S. ZHANG. Fusion of LBP and HOG using multiple kernel learning for infrared face recognition. In: *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. Wuhan: IEEE, 2017, pp. 81–84. ISBN 978-1-5090-5507-4. DOI: 10.1109/ICIS.2017.7959973.

[14] VIOLA, P. and M. J. JONES. Robust Real-Time Face Detection. *International Journal of Computer Vision*. 2004, vol. 57, iss. 2, pp. 137–154. ISSN 1573-1405. DOI: 10.1023/B:VISI.0000013087.49260.fb.

[15] WANG, Y.-Q. An Analysis of the Viola-Jones Face Detection Algorithm. *Image Processing On Line*. 2014, vol. 4, iss. 1, pp. 128–148. ISSN 2105-1232. DOI: 10.5201/ipol.2014.104.

[16] TAN, H., B. YANG and Z. MA. Face recognition based on the fusion of global and local HOG features of face images. *IET Computer Vision*.

2014, vol. 8, iss. 3, pp. 224–234. ISSN 1751-9640. DOI: 10.1049/iet-cvi.2012.0302.

[17] CAIFENG, S. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*. 2012, vol. 33, iss. 4, pp. 431–437. ISSN 0167-8655. DOI: 10.1016/j.patrec.2011.05.016.

[18] DALAL, N. and B. TRIGGS. Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. San Diego: IEEE, 2005, pp. 886–893. ISBN 0-7695-2372-2. DOI: 10.1109/CVPR.2005.177.

[19] HEATON, J. *Introduction to Neural Networks for Java*. 2nd ed. St. Louis: Heaton Research, Inc., 2008. ISBN 978-1-604-39008-7.

[20] EL-ABED, M. and C. CHARRIER. Evaluation of Biometric Systems. *New Trends and Developments in Biometrics*. 1st ed. New York: InTech, 2012. ISBN 978-953-51-0859-7.

[21] MARTINEZ, A. and R. BENAVENTE. The AR Face Database. In: *CVC Technical Report #24* [online]. 1998. `http://www2.ece.ohio-state.edu/~aleix/`.

## About Authors

**Jaromir TOVAREK** was born in 1989 in Moravska Trebova. He received M.Sc. degree in telecommunications from VSB–Technical University of Ostrava, Czech Republic, in 2013 and he continues in studying Ph.D. degree at the same university. His research is focused on signal processing, face recognition, speaker recognition and speech recognition.

**Miroslav VOZNAK** is an associate professor with Department of Telecommunications, the department chair in Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava, Czech Republic. He received his Ph.D. degree in telecommunications in 2002 and topics of his research include next generation networks, IP telephony, speech quality and network security. He is a senior member of IEEE Communications Society and many boards of conferences supported by IEEE such as TSP, INBIS or CN.

**Jan ROZHON** was born in 1986. He received his Ph.D. degree in telecommunications from VSB–Technical University of Ostrava, Czech Republic, in 2015. He is currently employed at the Department of Telecommunications. His research is focused on VoIP technology, namely performance of SIP network elements. He is with CESNET since 2009 as a researcher.

**Filip REZAC** was born in 1985. He received his Ph.D. degree in telecommunications from VSB–Technical University of Ostrava, Czech Republic, in 2015. He is currently employed at the Department of Telecommunications. His research is focused on IP telephony, computer networks and network security. He is with CESNET since 2009 as a researcher.

**Jakub SAFARIK** received his Ph.D. degree in telecommunications from VSB–Technical University of Ostrava, Czech Republic, in 2017. He is currently employed at the Department of Telecommunications. His research is focused on IP telephony, computer networks and network security. He is with CESNET as a researcher since 2011.

**Pavol PARTILA** received his Ph.D. degree in telecommunications from VSB–Technical University of Ostrava, Czech Republic, in 2017. He is currently employed at the Department of Telecommunications. His research is focused on speech processing, speech quality and VoIP.